

Northumbria Research Link

Citation: Malard, Lucie, Anwar, Muhammad Zohaib, Jacobsen, Carsten S. and Pearce, David (2021) Influence of Spatial Scale on Structure of Soil Bacterial Communities across an Arctic Landscape. *Applied and Environmental Microbiology*, 87 (5). e02220-20. ISSN 0099-2240

Published by: American Society for Microbiology

URL: <http://doi.org/10.1128/AEM.02220-20> <<http://doi.org/10.1128/AEM.02220-20>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/45116/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria
University**
NEWCASTLE



UniversityLibrary

1 Title: Spatial scale structure soil bacterial communities across an Arctic landscape

2 Running Title: Spatial scale structure Arctic soil bacterial communities

3 Lucie A Malard^{a,1,#}, Muhammad Zohaib Anwar^{b,2}, Carsten S Jacobsen^b and David A Pearce^{a,#}

4 Author affiliations:

5 ^a Faculty of Health and Life Sciences, Northumbria University, Newcastle-upon-Tyne NE1
6 8ST, United Kingdom

7 ^b Department of Environmental Sciences, Aarhus University, 4000 Roskilde, Denmark

8 ¹ present address: Department of Ecology and Evolution, University of Lausanne, 1015
9 Lausanne, Switzerland

10 ² present address: Department of Pathology and Laboratory Medicine, University of British
11 Columbia, 2211 Wesbrook Mall, Vancouver, British Columbia V6T 2B5, Canada

12 Corresponding Authors: #Lucie Malard: lucie.malard@unil.ch and #David Pearce:
13 david.pearce@northumbria.ac.uk

14

15 **Abstract**

16 Bacterial community composition is largely influenced by environmental factors, and this
17 applies to the Arctic region. However, little is known about the role of spatial factors in
18 structuring such communities. In this study, we evaluated the influence of spatial scale on
19 bacterial community structure across an Arctic landscape. Our results showed that spatial
20 factors accounted for approximately 10 % of the variation at the landscape scale, equivalent
21 to observations across the whole Arctic region, suggesting that while the role and
22 magnitude of other processes involved in community structure may vary, the role of
23 dispersal may be stable globally in the region. We assessed dispersal limitation by
24 identifying the spatial autocorrelation distance, standing at approximately 60 m, which
25 would be required in order to obtain fully independent samples and may inform future
26 sampling strategies in the region. Finally, indicator taxa with strong statistical correlations
27 with environment variables were identified. However, we showed that these strong taxa-
28 environment associations may not always be reflected in the geographical distribution of
29 these taxa.

30 **Importance**

31 The significance of this study is threefold. It investigated the influence of spatial scale on the
32 soil bacterial community composition across a typical Arctic landscape and demonstrated
33 that conclusions reached when examining the influence of specific environmental variables
34 on bacterial community composition are dependent upon the spatial scales over which they
35 are investigated. This study identified a dispersal limitation (spatial autocorrelation) distance
36 of approximately 60 m, required to obtain samples with fully independent bacterial
37 communities, and therefore, should serve to inform future sampling strategies in the region

38 and potentially elsewhere. The work also showed that strong taxa-environment statistical
39 associations may not be reflected in the observed landscape distribution of the indicator
40 taxa.

41 **Introduction**

42 Significant spatial structuring of soil microorganisms has been demonstrated at micro [μm -
43 mm] (1), plot [cm - to few meters] (1), landscape [hundreds of meters] (2), regional [kms]
44 (3), national (4, 5), continental (6), and global scales (7-9). Hence, the scale of investigation
45 is a key parameter to take into account in studies of bacterial biogeography. Martiny et al.
46 (10) further demonstrated the importance of spatial scale on environmental factors
47 identified influencing community composition in temperate soils. They found key
48 environmental drivers differed across spatial scales - ammonia-oxidizing bacterial (AOB)
49 community composition was dependent on distance, moisture and vegetation cover at the
50 plot scale; however, at the regional scale, diversity was mainly influenced by water
51 temperature, air temperature and moisture while nitrate concentration and air temperature
52 were predominant at the continental scale. Finally, when considering all scales together,
53 overall key drivers were geographic distance, sediment moisture, air temperature and
54 vegetation cover. However, most biogeographical studies only investigate communities at
55 one spatial scale (see Griffiths et al. (4), Tedersoo et al. (7), Bahram et al. (9) for further
56 examples). The landscape scale (few hundred of meters to few kilometers) is considered
57 highly relevant for studies of bacterial distribution patterns as it is the scale of human
58 activities (at which agricultural practices and land management are integrated). Hence, the
59 majority of studies at that scale investigate human-impacted landscapes (See Bru et al. (3),
60 Dao (11), Constancias et al. (2), Palta et al. (12) and Neupane et al. (13) for further

61 examples) with only few studies describing Arctic communities from few meters to 3 km
62 (14-16).

63 The first aim of this study was to evaluate the influence of the spatial scale on bacterial
64 community structure [Fig. S1] across an Arctic landscape [Fig 1]. Indeed, while the role of
65 environmental parameters such as pH (17, 18), total organic content (TOC) (19), moisture
66 content (20) and C:N ratio (21) on community composition in the Arctic has been
67 demonstrated, much less is known about the influence of spatial parameters (19). However,
68 determining the influence of environmental factors on communities remains an essential
69 step to avoid overestimating the role of the spatial scale. In addition to providing a better
70 understanding of the environmental factors influencing community structure, investigating
71 multiple scales provides better knowledge of the spatial structure, which facilitates the
72 development of sampling strategies where samples are collected beyond the spatial
73 autocorrelation distance and are, therefore, truly independent (22). As autocorrelation
74 distances have been identified from μm to km (22-25), with the potential of nested scales of
75 variability (26), and site to site variation, no standardized protocol exists for soil sampling for
76 metabarcoding studies (27, 28). Therefore, the second aim was to determine the minimum
77 distance required to obtain independent soil samples in the region [Fig. S1], which may
78 inform future sampling strategies in the Arctic. Finally, the last aim was to identify indicator
79 taxa which were closely associated with environmental variables and map their spatial
80 distribution across the landscape [Fig. S1]. Previous studies have attempted to identify
81 indicator taxa that could be used for environmental monitoring (for example Simonin et al.
82 (29) and Yang et al. (30) in rivers or Hermans et al. (31) in soils). As indicator taxa (32)
83 highlight OTUs with strong environmental associations that may respond to ecological
84 change, we expected their distribution to closely follow that of environmental parameters.

85

86 **Results**87 ***Environmental factors***

88 Results showed that all 35 environmental variables had a significant impact on bacterial
89 community structure with approximately 73 % of the variance explained by environmental
90 factors [Table 1]. Overall, five key factors (TOC, pH, conductivity, aluminium and arsenic)
91 had the most influence on bacterial community dissimilarity explaining 30 % of variation in
92 total. While all other environmental factors individually explained between 0.9 % and 2.4 %
93 of the variation, the combined soil elemental composition (excluding pH, conductivity and
94 TOC) accounted for 51.5 % of the total variation in bacterial community composition.

95 ***Variation partitioning***

96 A total of 9 dbMEMs vectors were built using (x,y) geographic coordinates and after forward
97 selection, five dbMEMs were identified as significantly impacting bacterial community
98 diversity and used in subsequent analyses. The variation partitioning analysis differentiated
99 the effect of environmental factors, linear trend and spatial vectors on community
100 composition [Fig. 2]. The environmental fraction X1 explained 73 % of the variance [Table
101 S1], equal to the finding by the adonis function and confirming the success of the variation
102 partitioning analysis. Using the adjusted R^2 values only as they accounted for the number of
103 variables in the model, environmental factors explained 54 % of the variance, of which 38 %
104 were not spatially structured (fraction [a]). The spatial component (X2 + X3) explained 25.6
105 % of the variation, of which 16.3 % could be explained by induced spatial dependence. This
106 was illustrated by fractions [d], [f] and [g], which represented spatially structured

107 environmental variables where the spatial structure of these environmental variables
108 induced a similar spatial structure in the response data, highlighting the need to evaluate
109 the influence of the environment on communities. The remaining 9.3 % of the spatial
110 component represented spatial autocorrelation. The linear trend accounted for 3.8 % of the
111 variance (fraction [b]) while spatial vectors explained 5.5 % of the variation. Fraction [e] had
112 a negative R^2 and could be considered null, as prescribed in D. Borcard et al. (33). Each
113 fraction (X1, X2, X3) was tested individually and was significant (ANOVA, $p < 0.001$). In total,
114 62.8 % of the bacterial community dissimilarity could be explained by environmental and
115 spatial factors while the remaining 37.2 % of the variance could not be explained by the
116 variables measured in this study.

117 ***Spatial scale and autocorrelation***

118 The distance-decay curve illustrated the increase in community dissimilarity with increasing
119 distance [Fig. 3A]. The power model was better fitted ($R^2 = 0.2261$, $p = 0.005$) than the linear
120 regression ($R^2 = 0.1844$, $p < 2.2 \times 10^{-16}$). Spatial autocorrelation was visualised on the
121 distance-decay curve [Fig. 3], where geographically close communities were more similar up
122 to 60 m. This was illustrated with the power model on the distance-decay curve, where the
123 blue curve begins to plateau [Fig. 3A]. To further characterise the spatial autocorrelation
124 distance, a Mantel correlogram was used [Fig. 3B] to compute the Mantel statistic between
125 the geographic distance and bacterial community dissimilarity distance (Bray Curtis). The
126 spatial autocorrelation was positive for the first distance class of 21 m, indicating that the
127 bacterial communities were more similar than expected by chance. The second distance
128 class of 63 m displayed no spatial autocorrelation, indicating random distribution beyond 63

129 m. Other distance classes presented negative autocorrelations indicating that these
130 bacterial communities were more different than expected by chance.

131 Geography also had some influence on environmental conditions with sites closer together
132 being more similar. The spatial autocorrelation of environmental variables was first
133 visualised in figure 3C, where geographically close sites were geochemically similar within 25
134 m. However, beyond approximately 25 m, site equally close or far could present similar
135 environmental conditions, as illustrated by the autocorrelation distance [Fig. 3C]. This was
136 also illustrated by the weak linear regression ($R^2 = 0.019$, $p < 2.2 \times 10^{-16}$) and the best-fitted
137 power model ($R^2 = 0.087$, $p = 0.005$). Spatial autocorrelation was further tested for each
138 individual variable using the semi-variograms produced prior to kriging. As semi-variograms
139 are specific to each variable, the spatial autocorrelation distances were unique to each
140 parameter. All the semi-variograms produced prior to Kriging indicated positive
141 autocorrelations oscillating between 1 m and 100 m, depending on the variable tested,
142 further illustrating the importance of the scale of investigation [Fig. S2].

143 ***Spatial distribution across the landscape***

144 Using an ordinary kriging method and after examining the semi-variograms, the spatial
145 distribution of alpha diversity and key environmental variables were mapped across the
146 landscape [Fig. 4]. The bacterial richness, diversity and evenness changed across the
147 landscape [Fig. 4(A, B & C)], and kriged maps illustrated the relationships between diversity,
148 evenness and richness. Overall, low richness indicated low diversity and low evenness,
149 further observed using linear models [Fig. S3]. The kriged maps of alpha diversity and
150 environmental variables showed the strong heterogeneity at the landscape scale with
151 changes from high to low concentrations within just a few meters [Fig. 4(D, E & F)].

Indicator taxa

The indicator species analysis identified 163 true specialists (statistic >0.98) OTUs associated with 12 environmental variables. Indicator taxa were generally associated with the highest concentration of each element. The phylogenetic tree specific to indicator taxa illustrated the high taxonomic diversity of indicator taxa [Fig. 5], however, figure 6 demonstrated that identified indicator taxa do not necessarily follow environmental gradients as they are expected to. Of the four key factors (excluding pH) influencing bacterial communities [Table 1], only conductivity and arsenic had some indicator taxa associated. Indicators of conductivity (Cond) were restricted to two OTUs associated with high conductivity, both Bacteroidetes classified in the Cytophagales order [Fig. 5]. Peaks of high conductivity were visualised in figure 6A and correlated with peaks in abundance of the two OTUs identified [Fig. 6B, C]. Indicators of arsenic (As) were closely associated with barium (Ba) and were taxonomically diverse, with the majority classified as Actinobacteria, Alphaproteobacteria, Chloroflexi, Halanaerobiales and Firmicutes [Fig. 5]. Arsenic concentration appeared more homogeneous across the landscape [Fig. 6D] with an average concentration = 13 ppm, min = 1.81 ppm, max = 20.51 ppm. These indicator taxa of arsenic were all associated with high concentrations [Fig. 6E, F, G, H & I] and were also associated with high concentrations of barium in the soil. Iron (Fe) and manganese (Mn) are both essential elements of soils. Iron concentration was highly heterogeneous across the landscape, with a strong peak in concentration at one site [Fig. 6J]. This peak was reflected by the presence of unique indicator taxa of which the abundance was closely related to this high concentration [Fig. 6K, L]. Indicators of iron were diverse, with a large number of Proteobacteria (Alpha, Beta, Gamma), Chloroflexi, Bacteroidetes, Cyanobacteria, Planctomycetes and Verrucomicrobia [Fig. 5]. On the other hand, manganese concentration was heterogeneous across the

176 landscape [Fig. 6M] but unlike other indicator taxa, they were associated with low
177 concentrations in the soil [Fig. 6N, O]. The indicator taxa of manganese were predominantly
178 classified as Proteobacteria [Fig. 5] and were also closely related to low concentrations of
179 niobium (Nb), lead (Pb) and zirconium (Zr), however, they were associated with high
180 concentrations of molybdenum (Mo). Indicator taxa of strontium (Sr) were limited to five
181 OTUs, an unknown Verrucomicrobium, a *Ca. Saccharibacterium* (TM7), a
182 Deltaproteobacterium and two Alphaproteobacteria while indicators of zinc (Zn) were
183 classified in all almost all phyla [Fig. 5], illustrating the wide array of specialist taxa
184 associated with high concentrations of zinc.

185 Discussion

186 *Key environmental factors influencing bacterial communities*

187 Total organic carbon, pH and conductivity were identified as the key drivers of bacterial
188 diversity across the Arctic landscape and are also commonly identified in studies across the
189 globe (8, 34-38). While pH was previously identified as the primary driver of bacterial
190 diversity in Arctic soils across the whole Arctic region (19); here, at the landscape scale, TOC
191 was identified as the primary factor influencing bacterial community structure and was
192 tightly linked with soil moisture. Generally, soil organic carbon content increases with
193 increasing precipitation and decreasing temperature (39). In the Arctic tundra, not only
194 precipitation but snowmelt and permafrost thaw have major impacts on soil moisture and
195 hydrology across the landscape (40, 41). In this study, where pH was on average
196 acidoneutral at 6.05 ± 0.36 with very few acidic patches, but organic carbon content was
197 very patchy (6 % - 46 %); the role of TOC in bacterial community structure is perhaps not

198 surprising. However, it highlights the importance of investigating different spatial scales as
199 drivers at the global scale may not necessarily be the same across the landscape of interest.

200 Aluminium and arsenic were the fourth and fifth environmental variables accounting for the
201 most variation in bacterial community structure [Table 1]. Aluminium is one of the most
202 abundant metal in the Earth's crust and microorganisms continuously interact with
203 aluminium in soils (42, 43). While aluminium lacks apparent biological function (42), the
204 aluminium ion (Al^{3+}) can be toxic to living organisms and is a function of the soil pH; the
205 concentration of toxic Al^{3+} gradually increases as pH decreases from pH = 6.2 (42, 43).
206 Here, little pH changes but large aluminium concentration variation were observed across
207 the landscape, which were not correlated to each other (linear regression: $R^2 = 0.00069$, $p =$
208 0.81). The toxicity of Al^{3+} may be influencing the bacterial community structure, however,
209 the concentration of Al^{3+} ions was not measured.

210 Arsenic is ubiquitous in low abundance in the natural environment and recognised as one of
211 the most toxic elements (44, 45). Here, a decrease in diversity and richness was observed
212 with increasing arsenic concentrations, which likely reflects the toxic effect of oxyanions of
213 arsenate on many bacteria, although some can use it as a terminal electron acceptor (44).
214 As with Al^{3+} , the chemical concentration of the various forms of arsenic was not measured
215 and therefore, cannot conclude that the toxicity has an influence on bacterial community
216 structure, although it is a possibility. Indicator taxa associated with high concentrations of
217 arsenic were diverse but dominated by Actinobacteria and Proteobacteria and was in
218 accordance with Dunivin et al. (45) who conducted a global survey of arsenic related genes
219 in soils and identified these phyla as harbouring more arsenic resistance genes.

220 All other elements measured had some influence on the observed bacterial community
221 [Table 1], from key major elements such as sulphur, calcium, silicon; to key trace elements
222 such as iron, manganese, magnesium, zinc, copper, molybdenum and cadmium; and other
223 elements, toxic or not, such as bromine, yttrium or lead. It should also be noted that while
224 TOC, pH and conductivity had a significant influence on bacterial community composition
225 (21.8 %), the soil elemental composition combined explained most of the variation (> 50 %).
226 This may serve to highlight the level of complexity of the factors influencing community
227 structure.

228 ***Indicator taxa***

229 Environmental variables were highly heterogenous across the landscape, which was
230 reflected by the distribution of alpha diversity and indicator taxa. The indicator species
231 analysis determined abundant OTU-environment associations and identified 163 OTUs that
232 could be considered true specialists in relation to 12 environmental variables. These OTUs
233 were generally associated with high concentrations of the variable in question except for
234 those associated with manganese, niobium, lead and zirconium which were representative
235 of low concentrations. As illustrated in the phylogenetic tree [Fig. 5], the diversity of these
236 indicator taxa was high, with numerous representatives of the Proteobacteria, Chloroflexi,
237 Bacteroidetes, Planctomycetes and Verrucomicrobia. The distribution of some indicator
238 taxa, selected for their reported relationship with the associated variable in the literature,
239 was mapped across the landscape to illustrate the association with the elements'
240 concentration. For arsenic, Clostridium and Clostridia-related (Halanaerobiales) taxa were
241 mapped as they have been identified with some role in arsenic cycling (44, 46) and with
242 arsenic-resistance genes (45). A Gemmatimonadetes and a Candidatus *Parcubacterium*

(clustered closely with the Cyanobacteria) were also mapped, as both have been identified with potential roles in arsenic cycling (46). The distribution of OTUs associated with iron were mapped and included a Cyanobacterium (47, 48) and a Deltaproteobacterium, a class with known taxa involved in iron cycling (47-49). Finally, the OTUs associated with manganese were also associated with other environmental variables and mainly identified as Proteobacteria. A Deltaproteobacterium and the only Chlamydiae identified were mapped, two classes associated with manganese cycling (48). While this analysis showed the strong associations of some OTUs with the measured environmental parameters, it also illustrated the difficulty of using indicator taxa for monitoring purposes due to the large number of associations identified and the high heterogeneity across the landscape. This was clear when the distribution of key indicator taxa was mapped across the landscape and did not clearly follow the distribution of the environmental variable associated. Furthermore, while indicator taxa may be identified, they do not necessarily participate in the associated element cycle. For instance, these OTUs may benefit from high concentration of arsenic due to higher tolerance to toxicity and decreased competition, without having any involvement in arsenic cycling.

Selection and dispersal structure bacterial communities

The variation partitioning analysis quantified the importance of both selection (deterministic) and dispersal (stochastic) on bacterial community structure. Environmental variables explained 54 % of the total variation, corresponding to selection and 16 % were spatially structured, corresponding to the induced spatial dependence. Then, spatial components (trend + dbMEMs) alone explained 10 % of the variation, illustrating spatial autocorrelation or dispersal (33). This is the same magnitude of influence as recorded in

266 Malard et al. (19) investigating biogeographical patterns across the whole Arctic region,
267 suggesting that the magnitude of influence of dispersal of bacterial community structure
268 may be stable in the Arctic.

269 More specifically, the distance-decay curve of environmental factors showed that edaphic
270 properties were spatially autocorrelated up to approximately 25 m, although this was the
271 overall spatial autocorrelation as each variable autocorrelated within different distances.
272 After that distance, environmental variables were independent, and this was illustrated by
273 the weak slope of the linear regression and the overall variability of edaphic properties. In
274 addition, even highly similar environmental conditions could harbour dissimilar bacterial
275 communities, further illustrating the potential role of dispersal and other processes such as
276 drift or diversification. The distance-decay curve of bacterial communities showed a positive
277 spatial autocorrelation distance at up to 60 m, which was further supported by the Mantel
278 correlogram. For the Arctic region as a whole, an autocorrelation distance within the same
279 order of magnitude, approximating 20 m, was previously identified (19). This limited
280 dispersal range in Arctic soils is in contrast with studies in other regions of the globe. For
281 instance, in a glacier forefield in southern Alaska, this distance was over 600 m (50) while in
282 British soils, it was below 1 km (4). It suggests that Arctic soil bacterial communities only
283 disperse to approximately 60 m and may form rather isolated island communities.
284 Therefore, the scale of sampling is important in these landscapes to capture community
285 variability and therefore, a minimum of 60 m should be maintained between sites to obtain
286 independent samples. Further investigations at other Arctic sites are required to determine
287 whether this applies across the whole Arctic region.

288 Overall, these results suggest that induced spatial dependence may be an important factor
289 shaping bacterial communities within 25 m, that is, as edaphic properties are very similar,
290 bacterial communities are also similar. Between 25 and 60 m, environmental variability
291 increased and yet, communities remained relatively similar, suggesting that dispersal may
292 be the primary process shaping bacterial communities. Beyond 60 m, the environment was
293 highly heterogeneous, bacterial communities were highly dissimilar and selection was likely
294 the main process structuring communities. While one process may dominate within each
295 distance category, it is still likely the combination of different processes (selection, dispersal,
296 diversification and drift) with different magnitudes still driving community assembly (51).

297 While 63 % of the variation (non-adjusted $R^2 = 81$ %) of bacterial community assemblage
298 could be explained, 37 % remained unexplained. Many factors, whether biotic or abiotic
299 could still be influencing bacterial communities. Based on the scale of this study, it is
300 unlikely that most climatic and topographic variables would have much influence on the
301 community structure variation. Instead, other edaphic factors such as total nitrogen or
302 phosphorus content, clay, silt and sand content but also the presence of ice or soil texture
303 may have more impact locally. Furthermore, biotic interactions such as competition and
304 predation within bacterial communities or with other members of the soil biota or higher
305 organisms may have some influence. For instance, grazing is one of the main disturbances to
306 the ecosystem locally, primarily by the Svalbard reindeer and the barnacle goose (52). In
307 addition to impacting the vegetation, they trample over the landscape and fertilise it and
308 therefore, grazing can have significant impacts on the ecosystem and has been shown to
309 decrease microbial respiration and the available carbon (53) while animal faeces increase
310 the available nitrogen and can increase bacterial abundance (54). Human presence may also
311 have some influence as the sampling site was close to another scientific research site with

312 open-top chambers, few cabins were located in the area, and the coal 'Mine 7' was still in
313 operation, approximately 1.5 km away and 400 m above the sampling site.

314 **Conclusion**

315 In this study, spatial factors accounted for approximately 10 % of the variation in community
316 composition at the landscape scale, equivalent to observations across the whole Arctic
317 region, suggesting that while the role and magnitude of other processes involved in
318 community structure may vary, the role of dispersal may be stable globally in the region.
319 Furthermore, the identification of different driving environmental factors at different scales
320 highlights their dependence upon the spatial scales over which they are investigated.
321 Overall, we suggest that induced spatial dependence may be shaping bacterial communities
322 within 25 m. Between 25 and 60 m, dispersal may be the primary process shaping bacterial
323 communities and beyond 60 m, selection is likely the main process structuring communities.
324 As dispersal may be limited to 60 m, and while further studies should be conducted, we
325 suggest that soil sampling in the region should be conducted beyond this distance to
326 capture landscape variability while collecting independent samples. Finally, by mapping the
327 spatial distribution of indicator taxa across the landscape, we showed that strong taxa-
328 environment statistical associations may not actually be reflected in the landscape
329 distribution of these bacterial taxa.

330 **Material and Methods**

331 ***Sampling site***

332 In July 2017, 44 soil samples were collected in Adventdalen, Svalbard [Fig. 1A] following the
333 sampling design depicted in figure 1B and characterised by 8 North-South transects of 5

334 samples each. Samples within each transect were approximately 50 m apart while the
335 distance between transects was approximately 100 m. On transect 6, extra samples were
336 collected 10 m and 1 m apart to investigate smaller scale patterns [Fig. 1B, 1C].

337 Adventdalen is a broad U-shaped valley open to the West, from which the mouth is located
338 approximately 2 km from Longyearbyen and 6 km from Svalbard Airport. Adventdalen was
339 deglaciaded about 10 ka BP (55) and permafrost is estimated to be 100 m thick close to the
340 shore. It is a typical Arctic landscape, in one of the driest areas of Svalbard, with an average
341 of 190 mm of annual precipitation, and mean annual temperature of -6 °C (56). The study
342 site was located approximately 9 km into the valley, 11 km away from Longyearbyen, at
343 78.17 °N, 16.02 °E. The vegetation is primarily dwarf shrub/grass heath, dominated by *Salix*
344 spp., mosses, lichens and *Graminea* spp. (57) [Fig. 1D]. The main disturbances to the site
345 come from grazing, primarily by the Svalbard reindeer (*Rangifer tarandus platyrhynchus*)
346 and the barnacle goose (*Branta leucopsis*) (52).

347 ***Sample collection and soil properties***

348 The coordinates from each site were recorded with a portable GPS. At each location, 50 g of
349 soil in the top 15 cm was collected using ethanol-cleaned trowels and Whirl-Pak bags
350 (Nasco, Fort Atkinson, WI, USA). Plant roots and rocks were removed manually in a class II
351 biological safety cabinet (ESCO, Singapore), samples were homogenised and frozen at -20 °C
352 before transportation to the United Kingdom for analyses. pH and conductivity were
353 measured in the laboratory in a 1:5 freshly thawed soil to water ratio, using a Mettler-
354 Toledo FE20 pH meter (Mettler-Toledo Instruments co., Shanghai, China) and a CMD500
355 conductivity meter (WPA, Cambridge, UK). Moisture content was measured gravimetrically
356 on soils after drying at 150 °C for 24 h and total organic content (TOC) was measured

357 gravimetrically by heating previously dried soils to 550 °C for 4 h. To analyse the elemental
358 composition of each sample, 5 g of thawed soil was placed in ceramic crucibles and left to
359 dry at 37 °C for 5 days. Dried samples were crushed to a fine powder using a mortar and
360 pestle, put in a powder sample cup, placed in the XRF spectrometer (X-Lab2000, Spectro,
361 Kleve, Germany) and analysed. Resulting concentrations were adjusted using calibrated
362 values and results were expressed in part per million (ppm).

363 ***DNA extraction and amplicon sequencing***

364 Soil DNA was extracted in duplicate for each sample using the PowerSoil kit (Qiagen, Hilden,
365 Germany) following the manufacturers' protocol. 16S rRNA gene libraries were constructed
366 using the universal primers 515F-806R (58) to amplify the V4 region. Amplicons were
367 generated using a high-fidelity Accuprime DNA polymerase (Invitrogen, Carlsbad, CA, USA),
368 were purified using AMPure magnetic bead capture kit (Agencourt, Beckman Coulter, MA,
369 USA) and quantified using a QuantIT PicoGreen fluorometric kit (Invitrogen). The purified
370 amplicons were then pooled in equimolar concentrations using a SequalPrep plate
371 normalization kit (Invitrogen), and the final concentration of the library was determined
372 using a SYBR green quantitative PCR (qPCR) assay. Libraries were mixed with Illumina-
373 generated PhiX control libraries and our own genomic libraries and denatured using fresh
374 NaOH. The resulting amplicons were sequenced on the Illumina MiSeq V2, 500 cycles.

375 ***Bioinformatics processing***

376 Raw paired-end reads were subjected to adaptor and primer clipping using Cutadapt (59)
377 resulting in $71,207 \pm 3,280$ reads per sample. Forward and reverse reads were merged using
378 FLASH (60). Reads with over 1.5 total expected errors per read and/or read length less than
379 245 base pairs were truncated during quality filtration using the Vsearch (61) filtering

380 module. It resulted in $64,917 \pm 4,291$ high quality merged reads per sample. Dereplication
381 was performed to identify unique sequences. A two-step chimera detection method was
382 used, first by aligning against ChimeraSlayer Gold database provided with SILVA (62), second
383 by using the *denovo* detection module. An open-reference operational taxonomic unit
384 (OTU) calling was performed on high-quality trimmed sequences at 97% similarity level
385 using the USEARCH (63) algorithm for clustering to generate operational taxonomical units
386 (OTUs). It resulted in (85 DNA samples) a total of 5,436,264 reads ($63,956 \pm 38,865$
387 reads/sample) assigned against 23,627 OTUs. Unique (chimera filtered) representative
388 sequences were aligned using the Python Nearest Alignment Space Termination (PyNAST)
389 (64) tool with a relaxed neighbour-joining tree built using FastTree (65). OTUs were assigned
390 taxonomy using the Lowest Common Ancestor (LCA) based Classification Resources for
391 Environmental Sequence Tags (CREST) (66) with a minimum classification score of 0.80
392 against SILVA release 128 as a reference database.

393 ***Data availability***

394 The sequencing dataset is deposited at the European Nucleotide Archive under the
395 BioProject accession PRJNA564217.

396 ***Statistical analysis***

397 Alpha diversity (richness, Shannon and Simpson indices) was calculated in QIIME v1.90 (67)
398 on the matrices resulting from multiple rarefactions with the smallest sample size (22316
399 reads) as maximum depth. Results were collated and averaged to obtain a single
400 representative value for each sample. The OTU table was normalised using cumulative-sum
401 scaling (CSS) (68). The resulting OTU table was input into R for subsequent analyses and the
402 Bray-Curtis dissimilarity distance was calculated using vegan (69).

403 To evaluate the environmental component, Pearson's correlation coefficients were
404 calculated using the corrplot package (70) to first identify possible correlations between
405 environmental variables. With these many variables, it was a necessary step to avoid
406 misinterpretation of the results (Katz, 2011). Coefficients over $|0.8|$ indicated strong
407 correlations [Fig. S4] and as such, variables were removed to keep only one representative
408 (Katz, 2011). For example, a high moisture content was correlated with a high TOC content
409 (Pearson's = 0.88), in this case, moisture was discarded as it is weather-dependent and is
410 expected to be more variable day to day than TOC. Of 48 parameters measured, 35 were
411 independent and considered to be representative. The distribution of the 35 remaining
412 environmental variables was investigated using the moments package (71) to assess the
413 skewness and kurtosis. Skewness evaluates the degree of distribution shift to one side or
414 another and a good distribution should equal 0, while kurtosis evaluates the tail distribution
415 and should also be close to 0 to assume normal distribution. Using diagnostic plots,
416 skewness and kurtosis, the necessary transformations to improve the unimodal distribution
417 of environmental variables were carried (summarised in Table S2) and collinearity was
418 verified again with Pearson's correlations [Fig. S5]. Transformed environment variables were
419 scaled and a sequential PERMANOVA was conducted using the adonis function
420 implemented in vegan with standard 999 permutations to identify environmental variables
421 significantly associated with the Bray-Curtis community dissimilarity.

422 To evaluate the spatial component, the geographic locations (x,y) of the sampling sites were
423 transformed to cartesian coordinates using the SoDA package (72) and the Euclidean
424 distance was calculated using vegan. Distance-decay curves were produced using linear
425 regressions of the Euclidean distance of the geographic locations against the Bray-Curtis
426 dissimilarity distance and the Euclidean distance of scaled environmental variables.

427 The presence of a linear trend (a systematic increase or decrease in the OTU data with (x,y)
428 coordinates) was visualised by the distance-decay curve [Fig. 3A] and tested by RDA and
429 ANOVA, as prescribed in Borcard et al. (33). As a significant linear trend was identified, the
430 OTU table was detrended by linear regression of the (x,y) coordinates. Distance-based
431 Moran's Eigenvector Maps (dbMEM) were constructed with (x,y) coordinates using the
432 adespatial R package (73). The significance of the spatial vectors (dbMEMs) was assessed
433 using the detrended OTU table and tested with ANOVA. Forward selection was conducted to
434 identify significant dbMEM vectors and the remaining dbMEMs were plotted using RDA.

435 Variation partitioning analysis (VPA) was used to assess the impact of environmental and
436 spatial factors on community composition (undetrended OTU table) and was conducted
437 using the environmental variables, (x,y) coordinates (linear trend) and significant dbMEM
438 vectors. Individual fractions were tested using RDA and ANOVA, as prescribed in Borcard et
439 al. (33).

440 To evaluate spatial autocorrelation, the detrended OTU table and the Euclidean distances of
441 cartesian coordinates (x,y) were used to produce a Mantel correlogram with standard 999
442 permutations using vegan. Semi-variograms were also produced using the autoKrig
443 function of the automap package (Hiemstra and Hiemstra, 2013) to use for geostatistical
444 analyses. Kriging was conducted using the autoKrig and automapPlot functions in the
445 automap package. Environmental variables and alpha diversity measures were interpolated
446 and mapped across the landscape.

447 Indicator taxa were determined by the Dufrene-Legendre indicator species analysis (32) to
448 identify OTUs that were specifically associated with different environmental variables. The
449 first step was to define categories for each environmental variable (i.e. high conductivity,

450 medium conductivity and low conductivity). To identify groups statistically rather than
451 subjectively, an automatic cluster approach was employed using the nbclust package (74),
452 which indicated the ideal number of groups (Table S2). Clusters were created using the
453 kmeans function (Table S2) and used with the multipatt function in the indicpecies package
454 with 999 permutations (32). Indicator taxa with a correlation statistic higher than 0.98 were
455 considered true specialists and used for subsequent analyses. The phylogenetic tree of
456 indicator taxa was built using the representative sequences from the identified indicator
457 taxa using FastTree method (65) and visualised using iTOL (75). Indicator taxa distribution
458 was mapped across the landscape by kriging, as previously described and Pearson
459 correlations were calculated between the indicator taxa and the environmental variables of
460 interest.

461 **Acknowledgements**

462 This work was supported by a grant from the European Commission's Marie Skłodowska
463 Curie Actions program under project number 675546. The authors also thank Edwin Sia and
464 UNIS for their participation and support in fieldwork. MiSeq sequencing of the 16S rRNA
465 gene was performed by the NU-OMICS sequencing service (Northumbria University).

466 LAM and DAP conceived and designed the study and sampling design. LAM carried the
467 fieldwork and laboratory work. MZA conducted the bioinformatics processing and LAM
468 conducted the statistical analysis. LAM drafted the manuscript and MZA, DAP and CSJ
469 revised and approved the final version.

470 **Conflict of interest**

471 The authors report no conflict of interests.

472 **References**

- 473 1. Nunan N, Wu K, Young IM, Crawford JW, Ritz K. 2003. Spatial distribution of bacterial
474 communities and their relationships with the micro-architecture of soil. *FEMS microbiology*
475 *ecology* 44:203-215.
- 476 2. Constancias F, Terrat S, Saby NP, Horrigue W, Villerd J, Guillemin JP, Biju-Duval L, Nowak V,
477 Dequiedt S, Ranjard L. 2015. Mapping and determinism of soil microbial community
478 distribution across an agricultural landscape. *MicrobiologyOpen* 4:505-517.
- 479 3. Bru D, Ramette A, Saby N, Dequiedt S, Ranjard L, Jolivet C, Arrouays D, Philippot L. 2011.
480 Determinants of the distribution of nitrogen-cycling microbial communities at the landscape
481 scale. *The ISME journal* 5:532.
- 482 4. Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. 2011. The bacterial
483 biogeography of British soils. *Environmental microbiology* 13:1642-1654.
- 484 5. Terrat S, Horrigue W, Dequiedt S, Saby NP, Lelièvre M, Nowak V, Tripied J, Regnier T, Jolivet
485 C, Arrouays D. 2017. Mapping and predictive variations of soil bacterial richness across
486 France. *PloS one* 12:e0186766.
- 487 6. Zhou J, Deng Y, Shen L, Wen C, Yan Q, Ning D, Qin Y, Xue K, Wu L, He Z. 2016. Temperature
488 mediates continental-scale diversity of microbes in forest soils. *Nature communications*
489 7:12083.
- 490 7. Tedersoo L, Bahram M, Pölme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios
491 AM, Thu PQ, Suija A. 2014. Global diversity and geography of soil fungi. *Science*
492 346:1256688.
- 493 8. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett
494 RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in
495 soil. *Science* 359:320-325.
- 496 9. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM,
497 Bengtsson-Palme J, Anslan S, Coelho LP, Harend H. 2018. Structure and function of the
498 global topsoil microbiome. *Nature* 560(7717):233-237.
- 499 10. Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC. 2011. Drivers of bacterial β -
500 diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*
501 108:7850-7854.
- 502 11. Dao TH. 2014. Landscape-scale geographic variations in microbial biomass and enzyme-labile
503 phosphorus in manure-amended Hapludults. *Biology and Fertility of Soils* 50:155-167.
- 504 12. Palta MM, Ehrenfeld JG, Giménez D, Groffman PM, Subroy V. 2016. Soil texture and water
505 retention as spatial predictors of denitrification in urban wetlands. *Soil Biology and*
506 *Biochemistry* 101:237-250.
- 507 13. Neupane S, Goyer C, Zebarth BJ, Li S, Whitney S. 2019. Soil bacterial communities exhibit
508 systematic spatial variation with landform across a commercial potato field. *Geoderma*
509 335:112-122.
- 510 14. Zak DR, Kling GW. 2006. Microbial community composition and function across an arctic
511 tundra landscape. *Ecology* 87:1659-1670.
- 512 15. Chu H, Neufeld JD, Walker VK, Grogan P. 2011. The influence of vegetation type on the
513 dominant soil bacteria, archaea, and fungi in a low Arctic tundra landscape. *Soil Science*
514 *Society of America Journal* 75:1756-1765.
- 515 16. Bottos EM, Kennedy DW, Romero EB, Fansler SJ, Brown JM, Bramer LM, Chu RK, Tfaily MM,
516 Jansson JK, Stegen JC. 2018. Dispersal limitation and thermodynamic constraints govern
517 spatial structure of permafrost microbial communities. *FEMS microbiology ecology*
518 94:fiy110.
- 519 17. Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. 2010. Soil bacterial diversity in
520 the Arctic is not fundamentally different from that found in other biomes. *Environmental*
521 *Microbiology* 12:2998-3006.

- 522 18. Siciliano SD, Palmer AS, Winsley T, Lamb E, Bissett A, Brown MV, van Dorst J, Ji M, Ferrari BC,
523 Grogan P. 2014. Soil fertility is associated with fungal and bacterial richness, whereas pH is
524 associated with community composition in polar soil microbial communities. *Soil Biology and*
525 *Biochemistry* 78:10-20.
- 526 19. Malard LA, Anwar MZ, Jacobsen CS, Pearce DA. 2019. Biogeographical patterns in soil
527 bacterial communities across the Arctic region. *FEMS microbiology ecology* 95.
- 528 20. Dimitriu PA, Grayston SJ. 2010. Relationship between soil properties and patterns of
529 bacterial β -diversity across reclaimed and natural boreal forest soils. *Microbial ecology*
530 59:563-573.
- 531 21. Docherty KM, Borton HM, Espinosa N, Gebhardt M, Gil-Loaiza J, Gutknecht JL, Maes PW,
532 Mott BM, Parnell JJ, Purdy G. 2015. Key edaphic properties largely explain temporal and
533 geographic variation in soil microbial communities across four biomes. *PloS one*
534 10:e0135352.
- 535 22. Katsalirou E, Deng S, Nofziger DL, Gerakis A, Fuhlendorf SD. 2010. Spatial structure of
536 microbial biomass and activity in prairie soil ecosystems. *European journal of soil biology*
537 46:181-189.
- 538 23. Dequiedt S, Saby N, Lelievre M, Jolivet C, Thioulouse J, Toutain B, Arrouays D, Bispo A,
539 Lemanceau P, Ranjard L. 2011. Biogeographical patterns of soil molecular microbial biomass
540 as influenced by soil characteristics and management. *Global Ecology and Biogeography*
541 20:641-652.
- 542 24. Bahram M, Kohout P, Anslan S, Harend H, Abarenkov K, Tedersoo L. 2016. Stochastic
543 distribution of small soil eukaryotes resulting from high dispersal and drift in a local
544 environment. *The ISME journal* 10:885-896.
- 545 25. Baker KL, Langenheder S, Nicol GW, Ricketts D, Killham K, Campbell CD, Prosser JI. 2009.
546 Environmental and spatial characterisation of bacterial community composition in soil to
547 inform sampling strategies. *Soil Biology and Biochemistry* 41:2292-2298.
- 548 26. Franklin RB, Mills AL. 2003. Multi-scale variation in spatial heterogeneity for microbial
549 community structure in an eastern Virginia agricultural field. *FEMS microbiology ecology*
550 44:335-346.
- 551 27. Taberlet P, Prud'homme SM, Campione E, Roy J, Miquel C, Shehzad W, Gielly L, Rioux D,
552 Choler P, Clement J-C. 2012. Soil sampling and isolation of extracellular DNA from large
553 amount of starting material suitable for metabarcoding studies. *Molecular ecology* 21:1816-
554 1820.
- 555 28. Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle
556 BE. 2019. DNA metabarcoding—Need for robust experimental designs to draw sound
557 ecological conclusions. *Molecular ecology* 28:1857-1862.
- 558 29. Simonin M, Voss KA, Hassett BA, Rocca JD, Wang SY, Bier RL, Violin CR, Wright JP, Bernhardt
559 ES. 2019. In search of microbial indicator taxa: shifts in stream bacterial communities along
560 an urbanization gradient. *Environmental microbiology* 21:3653-3668.
- 561 30. Yang Y, Li S, Gao Y, Chen Y, Zhan A. 2019. Environment-driven geographical distribution of
562 bacterial communities and identification of indicator taxa in Songhua River. *Ecological*
563 *Indicators* 101:62-70.
- 564 31. Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G. 2017. Bacteria as
565 emerging indicators of soil condition. *Applied and environmental microbiology* 83.
- 566 32. Cáceres MD, Legendre P. 2009. Associations between species and groups of sites: indices
567 and statistical inference. *Ecology* 90:3566-3574.
- 568 33. Borcard D, Gillet F, Legendre P. 2018. Numerical ecology with R. Springer.
- 569 34. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities.
570 *Proceedings of the National Academy of Sciences* 103:626-631.

- 571 35. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010.
572 Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME journal*
573 4:1340-1351.
- 574 36. Siles JA, Margesin R. 2016. Abundance and diversity of bacterial, archaeal, and fungal
575 communities along an altitudinal gradient in alpine forest soils: what are the driving factors?
576 *Microbial ecology* 72:207-220.
- 577 37. Wojcik R, Donhauser J, Frey B, Holm S, Holland A, Anesio AM, Pearce DA, Malard L, Wagner
578 D, Benning LG. 2018. Linkages between geochemistry and microbiology in a proglacial
579 terrain in the High Arctic. *Annals of Glaciology* 59:95-110.
- 580 38. Malard LA, Pearce DA. 2018. Microbial diversity and biogeography in Arctic soils.
581 *Environmental microbiology reports* 10:611-625.
- 582 39. Post WM, Emanuel WR, Zinke PJ, Stangenberger AG. 1982. Soil carbon pools and world life
583 zones. *Nature* 298:156.
- 584 40. Callaghan TV, Johansson M, Brown RD, Groisman PY, Labba N, Radionov V, Barry RG,
585 Bulygina ON, Essery RL, Frolov D. 2011. The changing face of Arctic snow cover: A synthesis
586 of observed and projected changes. *Ambio* 40:17-31.
- 587 41. Bonnaventure PP, Lamoureux SF. 2013. The active layer: A conceptual review of monitoring,
588 modelling techniques and changes in a warming climate. *Progress in Physical Geography*
589 37:352-376.
- 590 42. Piña RG, Cervantes C. 1996. Microbial interactions with aluminium. *Biometals* 9:311-316.
- 591 43. Barabasz W, Albinska D, Jaskowska M, Lipiec J. 2002. Ecotoxicology of aluminium. *Polish*
592 *journal of environmental studies* 11:199-204.
- 593 44. Oremland RS, Stolz JF. 2003. The ecology of arsenic. *Science* 300:939-944.
- 594 45. Dunivin TK, Yeh SY, Shade A. 2019. A global survey of arsenic-related genes in soil
595 microbiomes. *BMC biology* 17:45.
- 596 46. Jia Y, Huang H, Zhong M, Wang F-H, Zhang L-M, Zhu Y-G. 2013. Microbial arsenic methylation
597 in soil and rice rhizosphere. *Environmental science technology* 47:3141-3148.
- 598 47. Emerson D, Scott JJ, Benes J, Bowden WB. 2015. Microbial iron oxidation in the arctic tundra
599 and its implications for biogeochemical cycling. *Applied and Environmental Microbiology*
600 81:8066-8075.
- 601 48. Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. 2017. Metal availability and the
602 expanding network of microbial metabolisms in the Archaean eon. *Nature Geoscience*
603 10:629-636.
- 604 49. Lipson DA, Raab TK, Parker M, Kelley ST, Brislawn CJ, Jansson J. 2015. Changes in microbial
605 communities along redox gradients in polygonized Arctic wet tundra soils. *Environmental*
606 *microbiology reports* 7:649-657.
- 607 50. Darcy JL, King AJ, Gendron E, Schmidt SK. 2017. Spatial autocorrelation of microbial
608 communities atop a debris-covered glacier is evidence of a supraglacial chronosequence.
609 *FEMS Microbiology Ecology* 93.
- 610 51. Martiny JBH, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC,
611 Kane M, Krumins JA, Kuske CR. 2006. Microbial biogeography: putting microorganisms on
612 the map. *Nature Reviews Microbiology* 4:102-112.
- 613 52. Strebel D, Elberling B, Morgner E, Knicker HE, Cooper EJ. 2010. Cold-season soil respiration in
614 response to grazing and warming in High-Arctic Svalbard. *Polar Research* 29:46-57.
- 615 53. Stark S, Grellmann D. 2002. Soil microbial responses to herbivory in an arctic tundra heath at
616 two levels of nutrient availability. *Ecology* 83:2736-2744.
- 617 54. Sørensen LI, Mikola J, Kytöviita M-M, Olofsson J. 2009. Trampling and spatial heterogeneity
618 explain decomposer abundances in a sub-arctic grassland subjected to simulated reindeer
619 grazing. *Ecosystems* 12:830-842.

- 620 55. Mangerud J, Bolstad M, Elgersma A, Helliksen D, Landvik JY, Lønne I, Lycke AK, Salvigsen O,
621 Sandahl T, Svendsen JI. 1992. The last glacial maximum on Spitsbergen, Svalbard. *Quaternary*
622 *Research* 38:1-31.
- 623 56. Christiansen HH. 2005. Thermal regime of ice-wedge cracking in Adventdalen, Svalbard.
624 *Permafrost and Periglacial Processes* 16:87-98.
- 625 57. Mora C, Vieira G, Pina P, Lousada M, Christiansen HH. 2015. Land cover classification using
626 high-resolution aerial photography in adventdalen, svalbard. *Geografiska Annaler: Series A,*
627 *Physical Geography* 97:473-488.
- 628 58. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson
629 JK, Caporaso JG, Fuhrman JA. 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and
630 fungal internal transcribed spacer marker gene primers for microbial community surveys.
631 *Msystems* 1:e00009-15.
- 632 59. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
633 reads. *EMBnet journal* 17:10-12.
- 634 60. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve
635 genome assemblies. *Bioinformatics* 27:2957-2963.
- 636 61. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool
637 for metagenomics. *PeerJ* 4:e2584.
- 638 62. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a
639 comprehensive online resource for quality checked and aligned ribosomal RNA sequence
640 data compatible with ARB. *Nucleic acids research* 35:7188-7196.
- 641 63. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
642 26:2460-2461.
- 643 64. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2009. PyNAST: a
644 flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266-267.
- 645 65. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for
646 large alignments. *PloS one* 5:e9490.
- 647 66. Lanzén A, Jørgensen SL, Huson DH, Gorfer M, Grindhaug SH, Jonassen I, Øvreås L, Urich T.
648 2012. CREST—classification resources for environmental sequence tags. *PloS one* 7:e49334.
- 649 67. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena
650 AG, Goodrich JK, Gordon JI. 2010. QIIME allows analysis of high-throughput community
651 sequencing data. *Nature methods* 7:335.
- 652 68. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial
653 marker-gene surveys. *Nature methods* 10:1200.
- 654 69. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of*
655 *Vegetation Science* 14:927-930.
- 656 70. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. 2017. Package ‘corrplot’. *Statistician* 56:316-
657 324.
- 658 71. Komsta L, Novomestky F. 2015. Moments, cumulants, skewness, kurtosis and related tests. R
659 package version 14.
- 660 72. Chambers J. 2008. Software for data analysis: programming with R. Springer Science &
661 Business Media.
- 662 73. Dray S, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G, Legendre P,
663 Madi N, Wagner H. 2017. Adespatial: multivariate multiscale spatial analysis. R package
664 version 0.0-9.
- 665 74. Charrad M, Ghazzali N, Boiteau V, Niknafs A, Charrad MM. 2014. Package ‘nbclust’. *Journal*
666 *of statistical software* 61:1-36.
- 667 75. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and
668 annotation of phylogenetic and other trees. *Nucleic acids research* 44:W242-W245.

669

670

671 **Table 1:** The relative influence of environmental factors on bacterial community structure,
 672 calculated by PERMANOVA using the adonis function. * 0.05 > p > 0.01, ** 0.01 > p > 0.001,
 673 *** p < 0.001.

Variable	R ²	Pr(>F)	Variable	R ²	Pr(>F)	Variable	R ²	Pr(>F)
TOC	0.089	0.001***	Sr	0.018	0.002**	Th	0.013	0.005**
pH	0.070	0.001***	S	0.016	0.001***	Ag	0.012	0.007**
Cond	0.059	0.001***	Cu	0.015	0.001***	Mo	0.012	0.013*
Al	0.041	0.001***	Te	0.015	0.002**	Sb	0.012	0.010**
As	0.041	0.001***	Ba	0.014	0.003**	Cd	0.011	0.023*
Br	0.024	0.001***	In	0.014	0.002**	Ta	0.011	0.016*
La	0.022	0.001***	Nb	0.014	0.004**	Tl	0.011	0.021*
Y	0.021	0.002**	Nd	0.014	0.008**	Zr	0.011	0.012*
Ca	0.018	0.003**	Si	0.014	0.004**	Zn	0.010	0.031*
Cl	0.018	0.001***	Fe	0.013	0.002**	Ge	0.009	0.046*
Cs	0.018	0.001***	I	0.013	0.006**	Sn	0.009	0.036*
Pb	0.018	0.001***	Mn	0.013	0.009**	Residuals	0.269	N/A

674

675

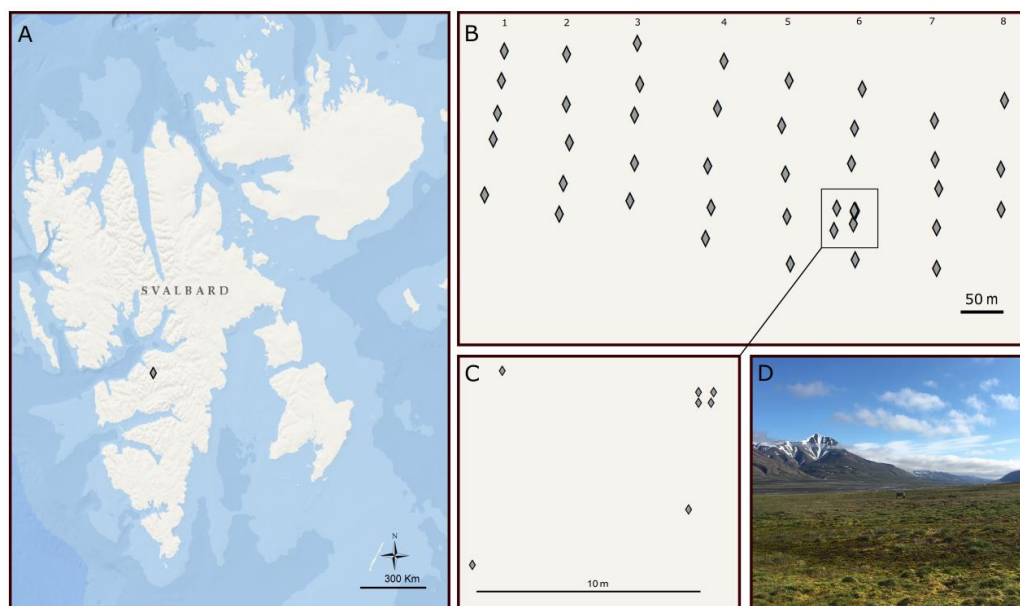
676

677

678

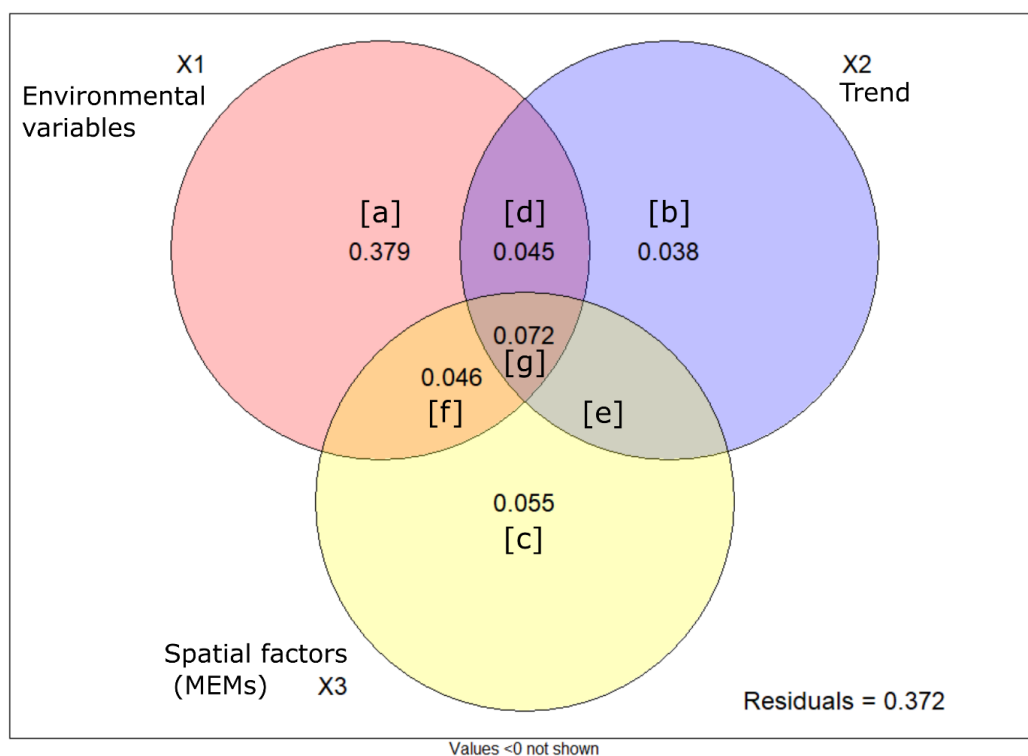
679 **Figures**

680



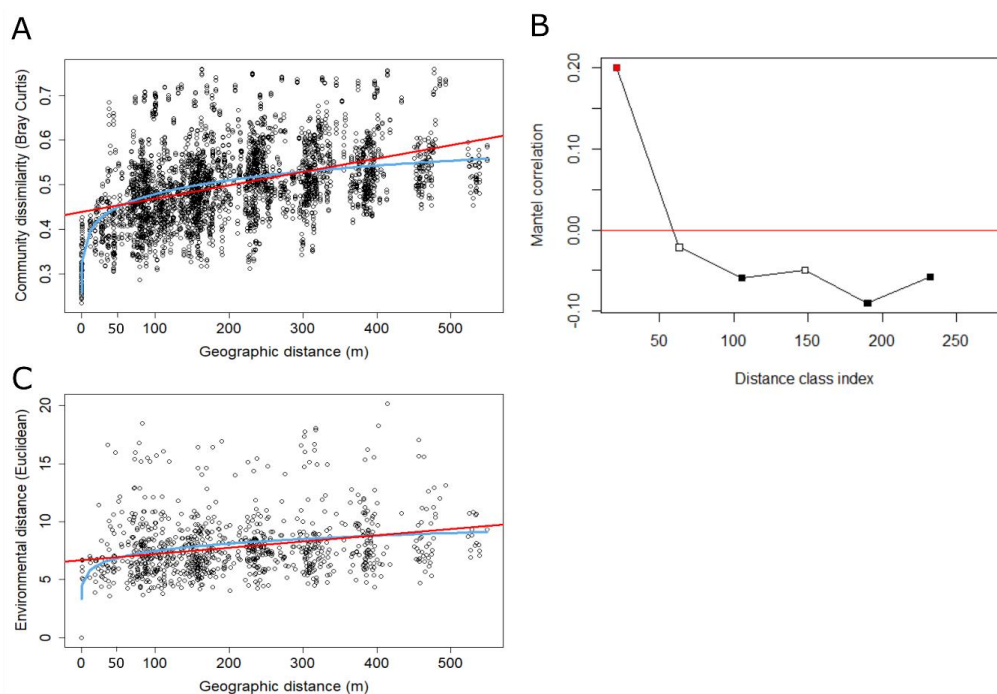
681 **Figure 1:** Map of sampling sites in (A) Svalbard. (B) Sampling design in 8 transects in
682 Adventdalen. (C) Smaller scale samples on transect 6. (D) View of Adventdalen.

683



684

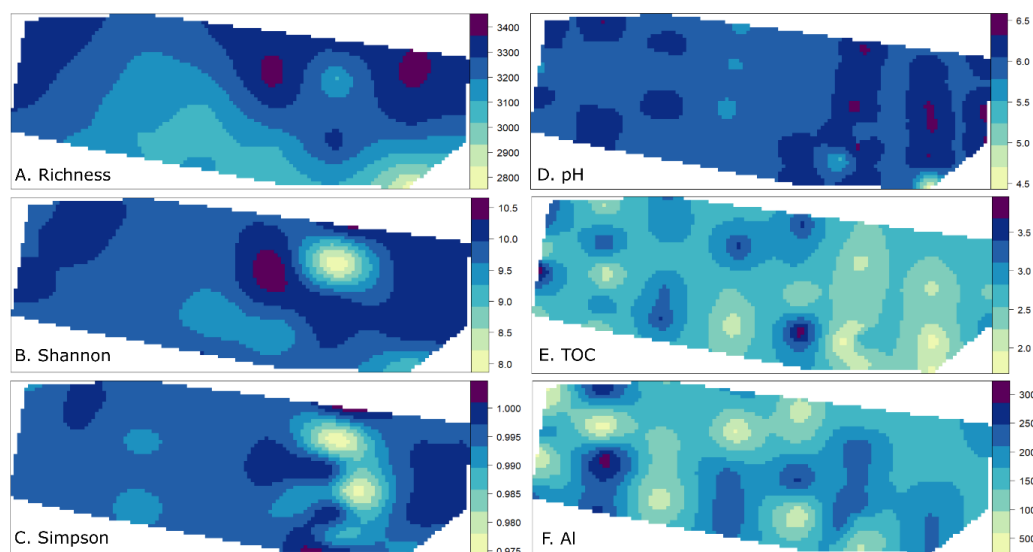
685 **Figure 2:** Venn diagram illustrating the results of the variation partitioning analysis on the
 686 influence of environmental variables and spatial factors on bacterial community
 687 composition. Results of each partition can be multiplied by 100 for the percentage of
 688 variation explained and are detailed in table S2.



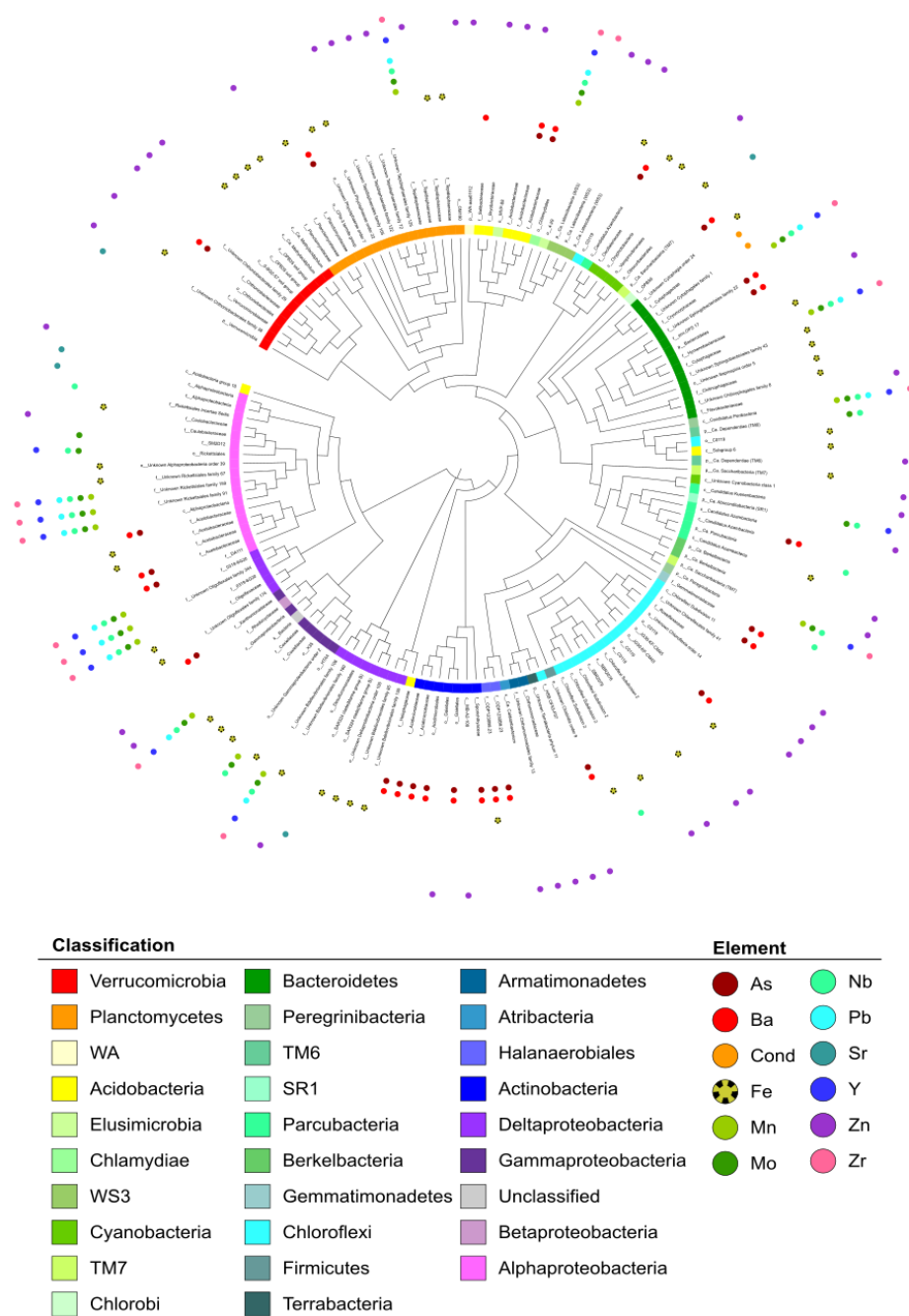
689

690 **Figure 3:** (A) Distance-decay curve illustrating the increase in bacterial community
 691 dissimilarity with increasing geographic distance. (B) Mantel correlogram of spatial
 692 autocorrelation illustrating the dispersal limitation. Red squares indicate positive significant
 693 autocorrelation which was only identified in the first distance class (0-21 m). Beyond 60 m,
 694 the autocorrelation was either negative (black squares) or not significant (white squares).
 695 (C) Distance-decay curve illustrating increasing environmental variation with increasing
 696 geographic distance. The red curve illustrates the linear regression and the blue curve is the
 697 power model.

698



699 **Figure 4:** Kriged maps of the spatial distribution across the landscape showing the
700 heterogeneity of (A) Richness, (B) Shannon index, (C) Simpson index, (D) pH, (E) Total
701 organic carbon and (F) Aluminium. The color bar of A, B, C indicates values of alpha diversity
702 while the color bar of environmental variables indicates element concentrations (see units of
703 each variable in table S2, taking into account data transformations).



704

705 **Figure 5:** Phylogenetic tree of indicator taxa associated with environmental variables
706 showing the high phylogenetic diversity. Coloured bands illustrate the taxonomy of each

OTU at the phylum level; labels indicate the taxonomy down to the family level if available.
Coloured points indicate the element associated.

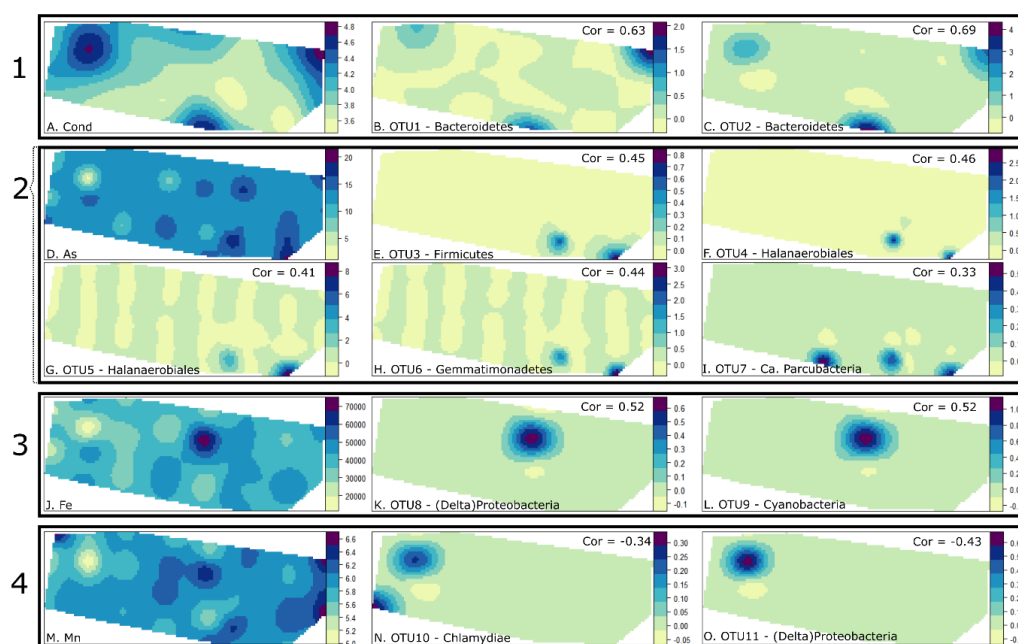
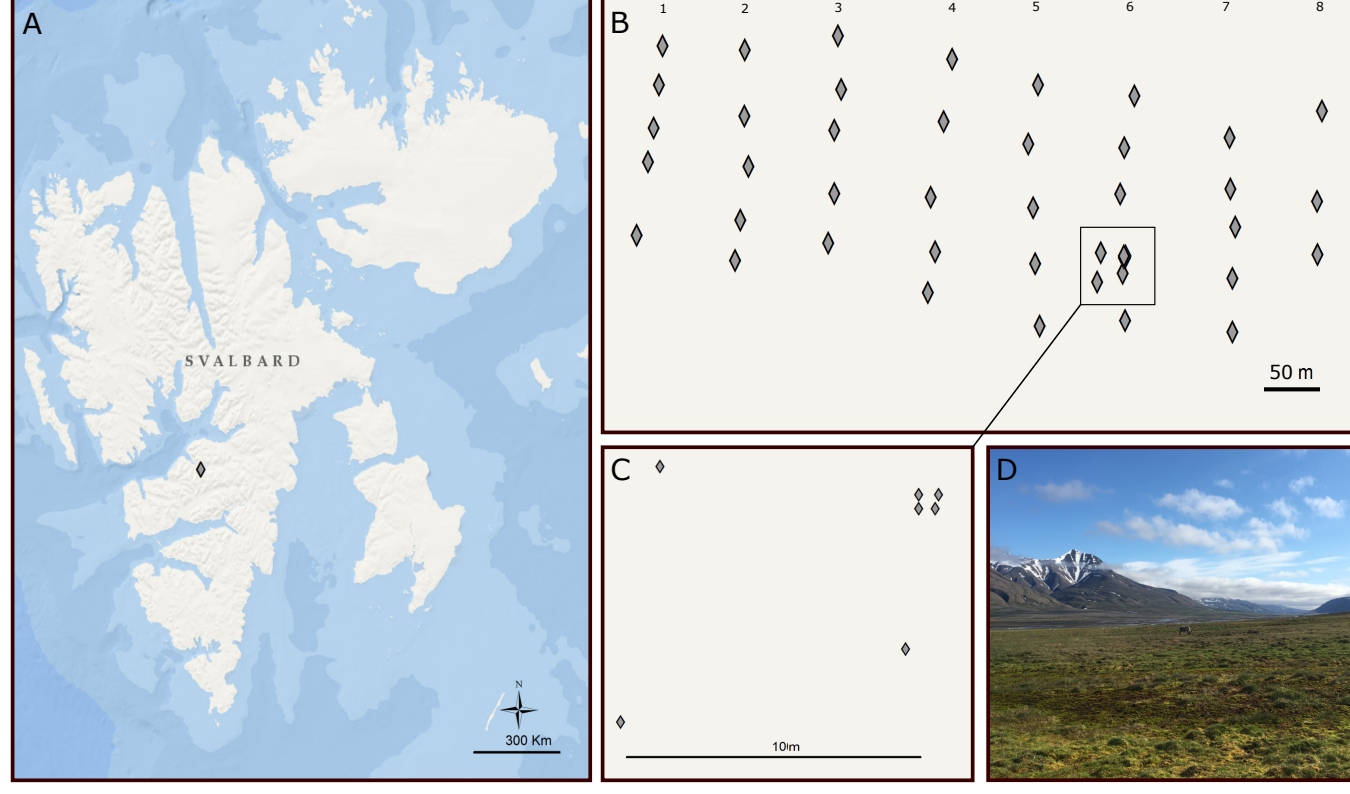
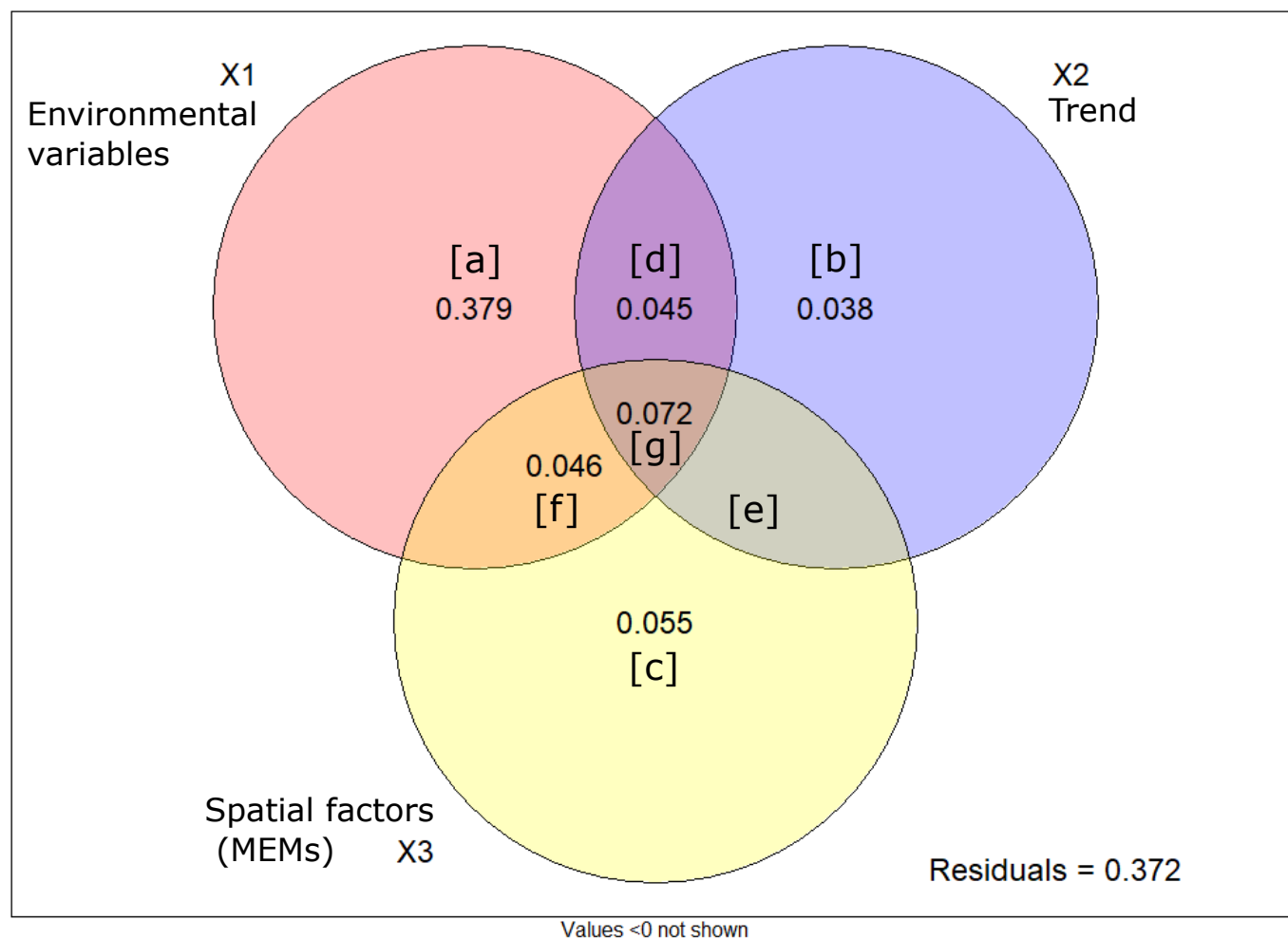
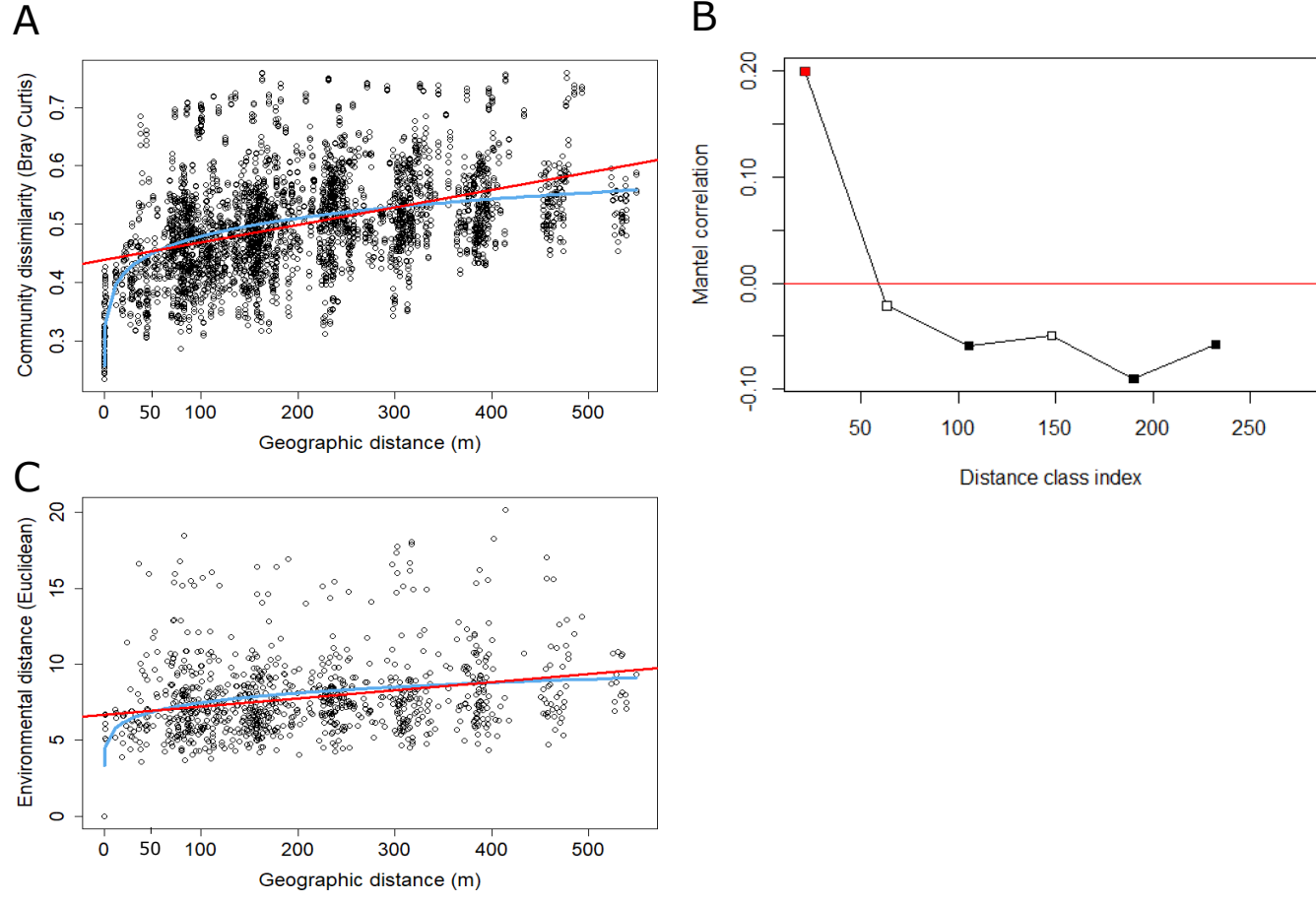
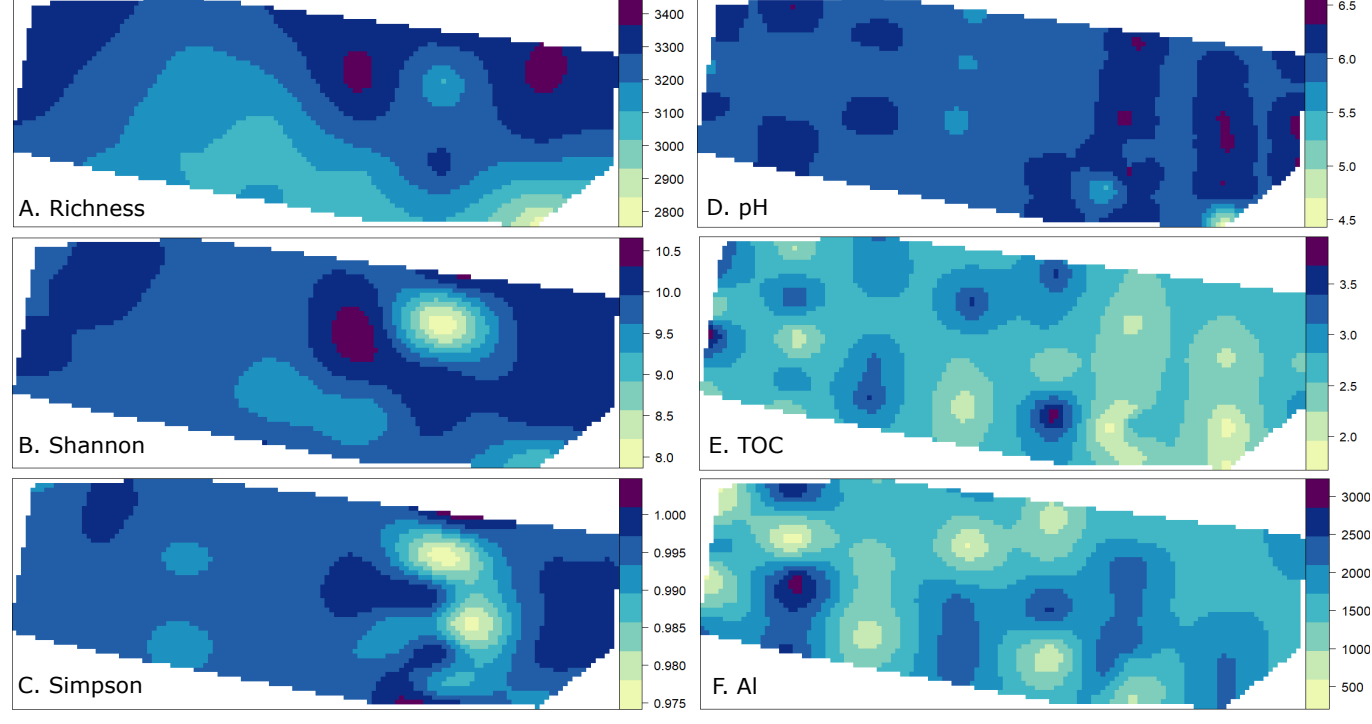


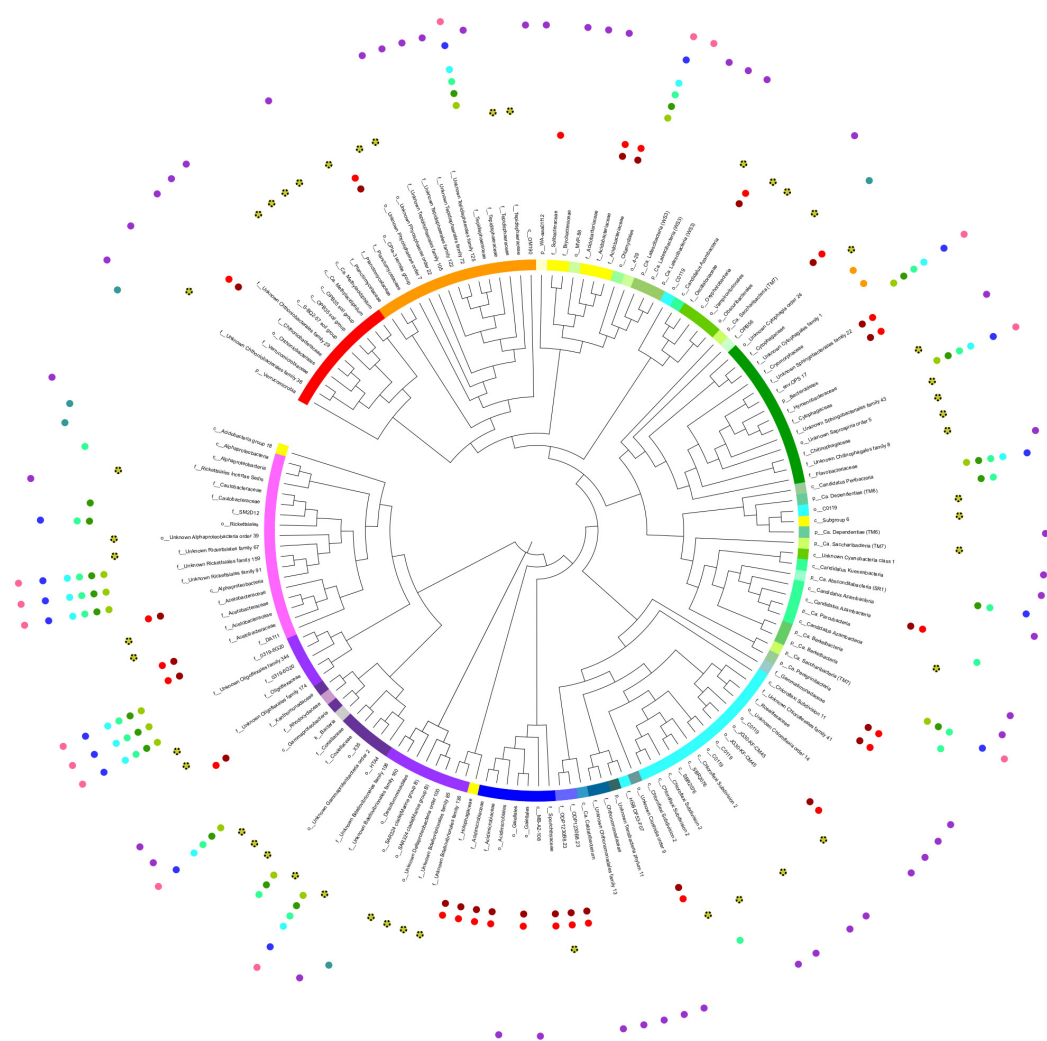
Figure 6: Spatial distribution across the landscape using Kriged map and illustrating the heterogeneous distribution. The color bar of environmental variables indicates element concentrations (Table S2 for units, considering data transformations) while the color bar for OTUs represents the relative abundance. Box 1:(A) Conductivity. (B) Phylum: Bacteroidetes, order: Cytophagales. (C) Phylum: Bacteroidetes, order: Cytophagales. Box 2:(D) Arsenic. (E) Phylum: Firmicutes, order: Unknown Clostridia. (F) Phylum: Halanaerobiales, order: Halanaerobiales. (G) Phylum: Halanaerobiales, order: Halanaerobiales. (H) Phylum: Gemmatimonadetes, +order: Gemmatimonadales. (I) Phylum: Ca. Parcubacteria, class: Ca. Azambacteria. Box 3:(J) Iron. (K) Phylum: Proteobacteria (Delta), order: Bdellovibrionales. (L) Phylum: Cyanobacteria, order: Oscillatoriaceae. Box 4:(M) Manganese. (N) Phylum: Chlamydiae, order: Chlamydiales. (O) Phylum: Proteobacteria (Delta), order: Oligoflexales.











Classification			Element	
Verrucomicrobia	Bacteroidetes	Armatimonadetes	As	Nb
Planctomycetes	Peregrinibacteria	Atribacteria	Ba	Pb
WA	TM6	Halanaerobiales	Cond	Sr
Acidobacteria	SR1	Actinobacteria	Fe	Y
Elusimicrobia	Parcubacteria	Deltaproteobacteria	Mn	Zn
Chlamydiae	Berkelbacteria	Gammaproteobacteria	Mo	Zr
WS3	Gemmatimonadetes	Unclassified		
Cyanobacteria	Chloroflexi	Betaproteobacteria		
TM7	Firmicutes	Alphaproteobacteria		
Chlorobi	Terrabacteria			

